

Ogg Vorbis SEBAGAI SALAH SATU ALTERNATIF METODE PEMAMPATAN SUARA MERUGI

Ogg Vorbis as an Alternatif Method for Lossy Sound Compression

Azis Wisnu Widhi Nugraha

asw_te@yahoo.com

Program Sarjana Teknik Unsoed Purwokerto

ABSTRACT

Multimedia technology has changed rapidly. How to get a small audio file with the high quality that still equal to the real quality is became the common issue in the sound compression world. Because the human ears can't hear all of sound component, it's allowed to compress an audio data using lossy method. A new lossy method to compress an audio data is Ogg Vorbis. This method has its own psychoacoustic model that made this method has the better sound quality if compared to the popular codec Mp3.

Keywords : sound compression, Ogg Vorbis, Ogg, Vorbis.

PENDAHULUAN

Perkembangan dunia multimedia dewasa ini yang cukup pesat dikarenakan kebutuhan manusia akan data multimedia cukup tinggi. Suara sebagai salah satu komponen multimedia memegang peranan yang cukup besar. Sejak ditemukan, teknik perekaman suara digital memiliki permasalahan dengan kapasitas media penyimpanan. Untuk menyimpan suara dengan kualitas CD audio dibutuhkan laju cuplikan (*sample rate*) sebesar 44,1 kHz, dengan kanal stereo (dua kanal : kiri dan kanan) dan jumlah bit kuantisasi untuk masing-masing kanal sebesar 16 bit (dua byte). Dengan demikian, dibutuhkan kapasitas media penyimpan sekitar 10MB/menit.

Ukuran data yang sedemikian besar menimbulkan beberapa permasalahan sebagai berikut.

- 1) Data suara tersebut akan membutuhkan kapasitas media penyimpanan yang besar.
- 2) Jika data tersebut akan ditransmisikan pada sebuah jaringan dibutuhkan lebar bidang (*bandwidth*) yang besar atau jika menggunakan lebar bidang yang kecil dibutuhkan waktu yang lama untuk mentransmisikannya.
- 3) Jika data tersebut akan ditransmisikan pada jaringan secara *real-time* maka

tidak boleh terjadi tunda waktu sehingga dibutuhkan lebar bidang yang sangat besar.

Dengan ukuran *file* data audio yang cukup besar, maka diperlukan suatu teknik pemampatan data audio untuk mengurangi ukuran data audio. Teknik pemampatan suara ini mengacu pada istilah *audio codecs*. Secara umum metode pemampatan suara dapat digolongkan dalam dua kelompok besar, yaitu metode pemampatan suara tak merugi dan metode pemampatan suara merugi.

Untuk memampatkan suara dimungkinkan dilakukan pemampatan secara merugi karena sesungguhnya telinga manusia memiliki keterbatasan untuk mendengarkan suara 'asli'. Bagaimana telinga manusia mendengarkan suara, melalui beberapa percobaan didekati dengan sebuah model yang disebut dengan *psychoacoustic model*. Model ini sesungguhnya berusaha untuk menentukan bagian mana dari suara yang akan terdengar oleh telinga manusia dan bagian mana yang tidak akan terdengar.

Salah satu metode pemampatan suara secara merugi yang masih terbilang baru adalah *Ogg Vorbis* yang dikembangkan oleh *Xiph.org* atau *Xiphopus Organization*. Teknik pemampatan suara ini dikembangkan oleh komunitas dengan lisensi model *BSD-style*. Dengan demikian, *Ogg Vorbis* merupakan salah satu jenis

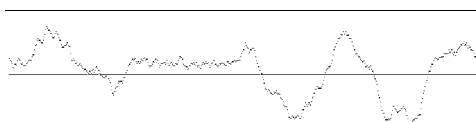
teknik pemampatan suara yang termasuk dalam kategori *free-software* (peranti lunak bebas). Jika dibandingkan dengan teknik pemampatan suara sekilas (Mp3), *Ogg Vorbis* memiliki beberapa keunggulan, seperti.

- 1) Kualitas subyektif *Ogg Vorbis* lebih baik dibandingkan dengan Mp3.
- 2) Dengan ukuran file yang sama dengan Mp3 akan diperoleh kualitas yang lebih baik atau dengan kata lain, dengan kualitas yang sama dengan Mp3 akan diperoleh ukuran file yang lebih kecil.

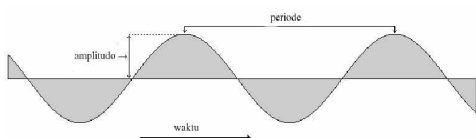
Pada tulisan ini akan sampaikan mengenai dasar pendigitalan suara, teknik pemampatan suara *Ogg Vorbis*, serta perbandingan hasil pemampatan suara menggunakan teknik Mp3 dan *Ogg Vorbis*.

PENDIGITALAN SUARA

Secara alamiah, sebuah gelombang suara memiliki bentuk yang sangat kompleks dan memiliki banyak sekali komponen frekuensi, sebagaimana dicontohkan pada Gambar 1. Meskipun sesungguhnya dimungkinkan sebuah gelombang suara merupakan gelombang sederhana yang hanya terdiri atas satu komponen frekuensi, sebagaimana terlihat pada Gambar 2. Bentuk gelombang suara yang kompleks sehingga memiliki komponen frekuensi yang sangat banyak ini mengakibatkan sulitnya merumuskan sebuah sinyal suara dalam sebuah persamaan sederhana.



Gambar 1 Contoh bentuk gelombang suara.

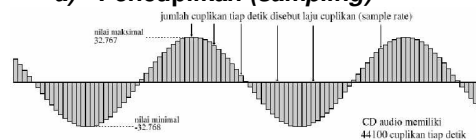


Gambar 2 Contoh gelombang sinus analog

Seiring dengan perkembangan teknik digital, dimungkinkan untuk merekam gelombang suara dalam bentuk digital. Untuk merubah data suara analog menjadi data suara digital digunakan teknik yang paling dasar yaitu *Pulse Code Modulation* (PCM). Adapun langkah-langkah

pendigitalan dengan menggunakan PCM adalah sebagai berikut (www.data-compression.com).

a) Pencuplikan (*sampling*)



Gambar 3 Contoh gelombang sinus tercuplik.

Pencuplikan adalah mengambil data amplitudo dari sebuah sinyal pada waktu tertentu secara berulang-ulang dengan periode tertentu (T detik). Jumlah cuplikan tiap detik disebut sebagai laju cuplikan (*sample rate*) yang besarnya sama dengan frekuensi pencuplik (f_s) atau $1/T$ Hz. Contoh gelombang sinus tercuplik dapat dilihat pada Gambar 3. Besar laju cuplikan minimal ditentukan oleh teorema *Nyquist* yang dirumuskan pada persamaan di bawah ini.

$$f_s = 2 \times f_{\max} \quad (1)$$

dimana :

f_s = frekuensi pencuplik (Hz)
 f_{\max} = frekuensi maksimal yang menjadi komponen frekuensi penyusun sinyal yang akan dicuplik (Hz)

b) Pembobotan (*quantization*)

Pembobotan/kuantisasi adalah pemberian nilai pada cuplikan yang diperoleh. Amplitudo minimal hingga maksimal data yang telah ditentukan dibagi dalam beberapa *interval* kuantisasi. Setiap *interval* diberi nilai. Dengan mengacu pada *interval* dan nilai tiap *interval* inilah data cuplikan akan dikuantisasi. Nilai hasil kuantisasi bukanlah nilai yang sesungguhnya, namun merupakan nilai pendekatan dari nilai yang sesungguhnya, sehingga akan terdapat perbedaan nilai hasil kuantisasi dengan nilai sesungguhnya yang disebut “derau kuantitasi”.

c) Penyandian (*coding*)

Langkah terakhir dari PCM adalah penyandian. Setiap nilai hasil kuantisasi akan sandikan dalam sandi biner. Pada prakteknya, langkah penyandian sangat berhubungan dengan langkah kuantisasi. Karena dengan menentukan banyaknya bit yang akan digunakan untuk penyandian, maka akan dapat ditentukan banyaknya *interval* kuantisasi. Sebagai contoh, dengan menggunakan tiga bit untuk penyandian maka akan diperoleh delapan *interval* kuantisasi, sedangkan jika dipergunakan 16

bit maka akan diperoleh 65.536 *interval* kuantisasi.

Adapun format CD audio yang merupakan format digital sinyal audio memiliki spesifikasi sebagai berikut :

- 1) laju cuplikan = 44,1 kHz
- 2) jumlah kanal = 2 kanal/
stereo (kiri dan kanan)
- 3) jumlah bit kuantisasi tiap kanal = 16 bit
= 2 *byte*.

PEMAMPATAN SUARA SECARA MERUGI

Sebagaimana telah disebutkan pada pendahuluan, *audio codecs* dapat dibagi dalam dua kelompok, yaitu pemampatan suara tidak merugi (*lossless compression*) dan pemampatan suara merugi (*lossy compression*) (Mitchell Graham, 2003). Pemampatan suara secara merugi akan mengakibatkan hilangnya beberapa komponen suara, sehingga suara yang dihasilkan dari data suara termampatkan hanyalah sesuatu yang “mirip” dengan data asli (Mitchell Graham, 2003; www.data-compression.com; www.cs.sfu.ca; M.H. Deghani dan Kuldip K.Paliwal, 2001).

Secara alamiah telinga manusia memiliki keterbatasan dalam mendengarkan suara. Dengan demikian memungkinkan dilakukan pemampatan suara secara merugi dengan cara membuang komponen-komponen suara yang “tidak didengar” atau yang “mungkin tidak didengar” oleh manusia. Untuk memutuskan apakah sebuah komponen suara tersebut didengar atau tidak oleh manusia dikenal istilah *psychoacoustic model*. Model ini mencoba memodelkan bagaimana telinga manusia “mendengar” suara. (www.cs.sfu.ca; *xiphopus mailing list archive*, 2003; Coleman Mike, 2001)

Secara garis besar proses pemampatan suara secara merugi dapat dilihat pada Gambar 4.

Dalam pemampatan suara merugi, data “mentah” berupa data PCM pertama-tama akan ditransformasikan ke dalam

kawasan frekuensi untuk mendapatkan sampel spektral dari sebuah data suara. Untuk melakukan transformasi dari kawasan waktu ke kawasan frekuensi ini biasanya digunakan teknik MDCT (*Modified Discrete Cosine Transform*).

Kemudian akan dilakukan analisa dengan menggunakan *psychoacoustic model* untuk menghilangkan komponen suara yang tidak diperlukan. Masing-masing *audio codecs* memiliki *psychoacoustic model*-nya masing-masing. Namun pada prinsipnya semua model tersebut berusaha untuk memodelkan “bagaimana telinga manusia mendengar”.

Setelah komponen suara yang sudah diperlukan dihilangkan akan dilakukan kuantisasi vektor dan penyandian *huffman* untuk mengurangi *redundancy* yang masih ada pada data, sehingga pada akhirnya akan diperoleh data termampatkan.

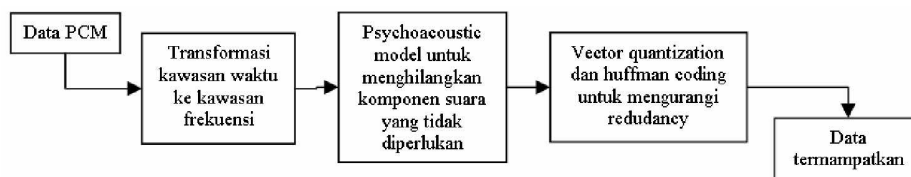
MODIFIED DISCRETE COSINE TRANSFORM (MDCT)

MDCT adalah transformasi bertumpuk orthogonal yang didasarkan pada ide *Time Domain Aliasing Cancellation* (TDAC). Sebelum data diolah dengan menggunakan MDCT, data dikelompokkan dalam blok – blok dengan menggunakan fungsi *windowing*. Di mana untuk MDCT, setiap *window* akan memiliki *overlap* sebesar 50% untuk lebar *window* yang sama dan akan lebih rumit lagi untuk lebar *window* yang berbeda. *Windowing* ini dimaksudkan untuk memperhalus hasil keluaran transformasi. (Sporer Th. et al, 1992; Wright Keith, 2003)

Jika diberikan *sample* dalam kawasan waktu dengan panjang blok *sample* sebesar *n sample*, dan $x_t(k)$ dengan $k = 0, 1, \dots, n-1$ adalah *sample* dalam kawasan waktu, serta $X_t(k)$ dengan $k = 0, 1, \dots, n/2-1$ adalah nilai *sample* pada kawasan frekuensi, maka persamaan MDCT adalah sebagai berikut.

$$X_t(m) = \sum_{k=0}^{n-1} x_t(k) \cos\left(\frac{\pi}{2} \left(2k+1 + \frac{n}{2}\right) (2m+1)\right)$$

$$m = 0, 1, \dots, \left(\frac{n}{2} - 1\right) \quad \dots (2)$$



Gambar 4 Teknik pemampatan suara merugi.

Sedangkan *Inverse MDCT* (IMDCT) adalah.

$$y_i(p) = f \left(\frac{1}{4} \sum_{m=0}^{\frac{n}{2}-1} X_i(m) \cos \left(\frac{\pi}{2} \left(2p+1 + \frac{n}{2} \right) (2m+1) \right) \right) \dots (3)$$

$$p = 0, 1, \dots, n-1$$

dimana $f(x)$ adalah fungsi *window* yang dapat berbeda-beda untuk masing-masing algoritma. Untuk ukuran blok frame yang sama, *window* yang digunakan simetris. Sebagai contoh, *codecs* Mp3, MPEG-2 dan AAC menggunakan fungsi *window* berikut.

$$w_k = \sin \left[\frac{\pi}{2} \left(k + \frac{1}{2} \right) \right] \dots (4)$$

Sedangkan untuk *Vorbis* digunakan fungsi *window* sebagai berikut.

$$w_k = \sin \left(\frac{\pi}{2} \sin^2 \left[\frac{\pi}{2} \left(k + \frac{1}{2} \right) \right] \right) \dots (5)$$

PSYCHOACOUSTIC MODEL

Pendengaran manusia memiliki keterbatasan dalam mendengarkan suara yang diterima. *Psychoacoustic model* akan berusaha mencari komponen suara mana yang terdengar dan membuang mana yang tidak terdengar. Contoh sederhana adalah telinga manusia hanya mampu mendengarkan suara antara frekuensi 20 Hz hingga 20 kHz, meski pada telinga dewasa biasanya frekuensi maksimal yang dapat di dengar adalah 16 kHz. Dengan demikian, komponen frekuensi di luar 20 Hz hingga 16 kHz dapat dibuang. Contoh lain adalah jika kita berada dalam lingkungan yang berisik, misalnya di dalam ruang mesin, kita tidak dapat mendengarkan suara siulan ringan, untuk itu kita dapat membuang komponen siulan ringan tersebut dari data, meski dalam prakteknya belum ada *psychoacoustic model* yang dapat melakukan hal tersebut.

Dalam membuat sebuah *psychoacoustic model* dikenalkan konsep-konsep berikut. (www.cs.sfu.ca)

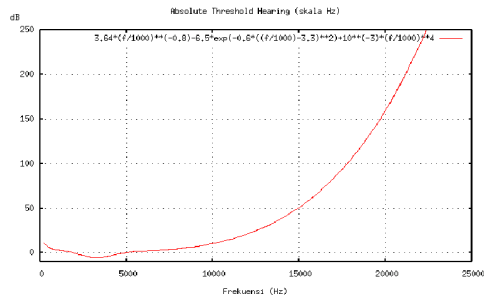
- 1) Ambang pendengaran mutlak/*Absolute Threshold of Hearing* (ATH)
Meskipun telinga manusia dapat mendengar frekuensi dari 20 Hz hingga 20 kHz, namun hal ini tidak berarti bahwa semua frekuensi didengar dengan cara yang sama. Pada frekuensi antara 2 hingga 4 kHz telinga manusia lebih sensitif dibandingkan dengan frekuensi yang lainnya. Sementara frekuensi

pembicaraan normal berkisar antara 500 Hz hingga 2 kHz.

Persamaan untuk ATH telinga manusia sebagai berikut.

$$ATH(f) = 3,64 \left(\frac{f}{1000} \right)^{-0,8} - 6,5 e^{-0,6 \left(\left(\frac{f}{1000} \right)^{-3,3} \right)^2} + 10^{-3} \left(\frac{f}{1000} \right)^4 (dB) \dots (6)$$

Persamaan di atas dapat digambarkan seperti tampak pada Gambar 5 di bawah ini.



Gambar 5 Ambang pendengaran mutlak manusia.

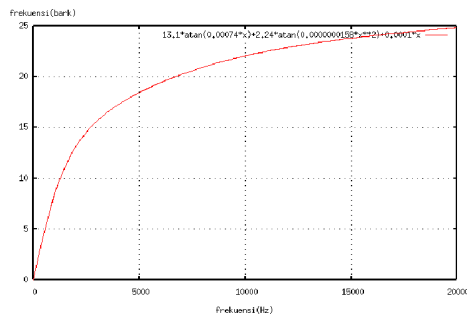
2) Frequency masking

Telinga manusia lebih mampu membedakan pola nada pada frekuensi rendah dibandingkan pada frekuensi tinggi. Sebagai contoh, telinga manusia akan lebih mudah membedakan frekuensi 500 Hz dan 600 Hz dibandingkan membedakan frekuensi 17 kHz dan 18 kHz. Hal ini berarti beberapa frekuensi yang berdekatan tidak dapat terdengar oleh telinga manusia jika muncul secara bersamaan.

Setelah melewati beberapa studi, diketahui bahwa antara 20 Hz hingga 20 kHz, frekuensi dapat dibagi dalam bandwidth kritis (*critical bandwidth*), yang tidak seragam, tidak linear, dan bergantung pada suara yang terdengar. Sinyal yang berada diluar satu *critical bandwidth* baru akan terdengar berbeda bagi telinga manusia. Lebar satu *critical bandwidth* disebut dengan 1 Bark.

Hubungan frekuensi dalam Hz dan *critical bandwidth* dalam Bark dalam Ogg Vorbis dipenuhi oleh persamaan berikut, digambarkan pada Gambar 6.

$$bark = 13,1 \arctan(0,00074 f) + 2,24 \arctan(0,0000000158 f^2) + 0,0001 f \dots (7)$$



Gambar 6 Hubungan antara frekuensi Hz dan frekuensi Bark pada Ogg Vorbis.

3) Temporal masking

Temporal masking adalah sebuah mekanisme pertahanan dari telinga yang diaktifkan untuk melindungi strukturnya yang halus dari suara yang keras. Saat menerima suara yang keras, telinga manusia akan bereaksi dengan berkontraksi sedikit, sementara akan mengurangi volume penerimaan suara yang mengikutinya. Reflek ini, seringkali disebut dengan "*blinking*" (dianalogikan dengan mata), ini berarti untuk melindungi struktur telinga yang halus dari daya suara yang merusak

LAJU BIT (BIT RATE)

Di muka telah disebutkan mengenai laju cuplikan (*sample rate*), yang berarti jumlah cuplikan dari data yang akan diperoleh tiap detiknya. Dalam dunia *digital audio* dikenal juga *bit rate* atau laju bit, yaitu jumlah bit yang digunakan untuk menyatakan data audio sepanjang satu detik. Sebuah data suara dengan kualitas CD audio yang tidak dimampatkan menggunakan memiliki laju bit sebagai berikut.

$$\begin{aligned} br_{cd} &= f_s \times n_q \times n_{ch} \\ br_{cd} &= 44,1kHz \times 16bit \times 2kanal \dots\dots\dots (8) \\ br_{cd} &= 1.411.200bps \\ br_{cd} &\approx 1.411kbps \end{aligned}$$

Pada teknik pemampatan suara merugi, dengan menggunakan laju bit sekitar 64 hingga 256 kbps kita sudah dapat menyimpan informasi yang "sama dengan" data suara sekualitas CD audio.

Permasalahan yang timbul adalah *bit rate* hanyalah berbicara mengenai ukuran dari data suara, tidak membicarakan mengenai kualitas dari suara tersebut. Sebagai contoh sebuah data suara yang menyimpan sebuah ketukan drum pada

suatu detik, ketika di mampatkan dengan *bit rate* sebesar 128 kbps kualitas suara yang dihasilkan sudah cukup memuaskan telinga pendengar bahkan mungkin sudah lebih dari cukup, karena sesungguhnya mungkin dengan laju bit sebesar 64 kbps kualitas suara pada detik tersebut sudah cukup memuaskan. Namun apabila pada detik yang lain musik berubah menjadi gaduh, pukulan drum menjadi sangat cepat dan mulai memukul *cymbal* dengan keras, *lead* gitar mulai bermain, *bass* gitar mulai dicabik dengan cepat, dan suara *vocal* melengking, dengan menggunakan *bit rate* sebesar 128 kbps maka suara pada detik tersebut akan terdengar lebih jelek. Karena sesungguhnya untuk mempertahankan kualitas suara pada detik tersebut harus menggunakan laju bit sebesar, misal 256 kbps atau bahkan lebih untuk menjaga kualitas suara yang dihasilkan.

Dengan demikian dalam pemampatan suara dikenal tiga jenis *bit rate*, yaitu. (Mitchell Graham, 2003)

1) Laju bit tetap/*Constant Bit Rate* (CBR)

Pemampatan suara menggunakan CBR, sebagaimana namanya, menggunakan laju bit yang tetap sepanjang data yang ada. Kekurangan dari CBR ini sebagaimana diilustrasikan di atas, namun keunggulan dari CBR kita dapat memperkirakan besar media penyimpanan yang dibutuhkan. Sebagai contoh jika sebuah *file* data audio sepanjang lima menit akan dimampatkan dengan kecepatan 256 kbps, maka besar media penyimpan yang dibutuhkan sebesar: 256 kbps x 300 detik : 8 bit = 9.600 kB = 9,6 MB.

2) Laju bit rata-rata/*Average Bit Rate* (ABR)

Proses *encoding* dengan menggunakan *average bit rate* (AVR) adalah proses *encoding* di mana *encoder* akan menggunakan *bit rate* rata-rata yang kemudian akan dikurangi atau ditambah sesuai dengan kebutuhan, namun masih dipertahankan nilai rata-rata dari *bit rate* yang digunakan. Pada umumnya *encoder* Mp3 menggunakan ABR.

3) Laju bit berubah/*Variable Bit Rate* (VBR)

Variable bit rate (VBR) adalah proses *encoding* di mana kualitas suara dari data dipertahankan sama sepanjang data. Dengan demikian tidak ada

patokan berapa besar *bit rate* yang akan digunakan. Besarnya *bit rate* dapat berubah – ubah sesuai dengan kebutuhan. Besarnya *bit rate* yang digunakan pada satu waktu tertentu diupayakan dapat mempertahankan kualitas suara yang akan dihasilkan pada saat proses *decoding*, hal ini bergantung pada *encoder* yang digunakan, dalam hal ini adalah *psychoacoustic model* yang digunakan oleh *codecs*. Meskipun kualitas dari data yang dihasilkan akan dapat dipertahankan dari awal hingga akhir, namun kelemahan dari teknik ini adalah besarnya data yang dihasilkan tidak dapat diperkirakan sebelumnya.

CODECS *Ogg Vorbis*

Encoder *Ogg Vorbis*

Dari Gambar 7, pertama-tama data PCM akan menjadi masukan bagi blok *windowing*, *Ogg Vorbis* menggunakan panjang *window* antara 64 hingga 8192 cuplikan (merupakan perpangkatan dari dua) yang saling bertumpangan sebesar 50%. Setelah dilakukan proses *windowing*, maka selanjutnya data cuplikan ditransformasikan dari kawasan waktu ke kawasan frekuensi dengan menggunakan MDCT (*Modified Discrete Cosine Transform*).

Setelah data ditransformasikan ke kawasan frekuensi, kemudian model *psychoacoustic* diterapkan pada data. Pada tahap inilah proses 'pembuangan' bagian yang tidak dapat terdengar oleh telinga manusia dilakukan. Proses ini pula yang akan menentukan besarnya rasio pemampatan. Keluaran dari blok ini berupa data *floor* yang merupakan representasi dari model resolusi rendah dari spektrum audio. *Ogg Vorbis* memiliki dua jenis data *floor*, yaitu '*floor* tipe 0' dan '*floor* tipe 1' (www.xiph.org). '*Floor* tipe 0' menggunakan LSP (*Line Spectral Pair*) untuk merepresentasikan model spektrum audio,

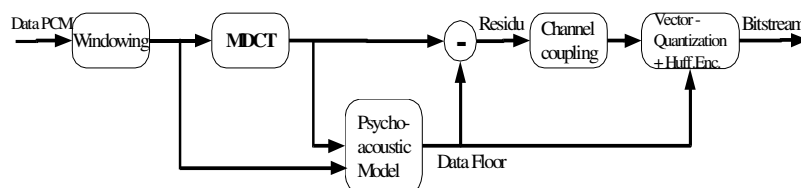
sedangkan '*floor* tipe 1' menggunakan algoritma interpolasi linear. Kemudian *floor* data akan dikurangkan dari data keluaran MDCT untuk menghasilkan data *residue* yang merupakan representasi struktur halus spectrum sinyal audio.

Data *residue* ini kemudian akan dimasukkan dalam blok kopling kanal untuk mengurangi *redundancy* dari kanal. Secara teoritis, *Ogg Vorbis* dapat memiliki kanal sebanyak 255 kanal (www.xiph.org), namun hingga saat ini kopling kanal lebih dari dua (stereo) belum didukung. Terakhir data *residue* dan *floor* disandikan dengan menggunakan kuantisasi vektor (VQ) dan sandi Huffman untuk meminimalkan *redundancy*. Keluaran blok ini merupakan data audio *Vorbis* yang disebut dengan data analisis. Data analisis *Vorbis* ini kemudian akan diberi *header* yang terdiri atas tiga macam *header* (*parameter header*, *comment header*, dan *codebooks header*). Data analisis yang telah diberi *header* ini disebut dengan data mentah *Vorbis*.

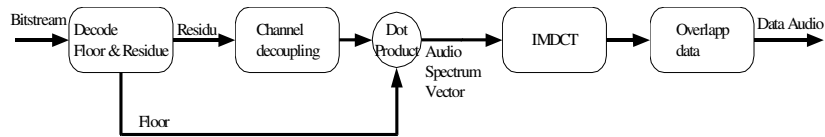
Data mentah *Vorbis* ini dapat langsung digunakan untuk pengiriman data menggunakan protokol yang telah memiliki proses sinkronisasinya sendiri seperti UDP/*Uni Datagram Protocol* (RTP/*Real Time Protocol*). Sedangkan untuk keperluan penyimpanan atau trasport seperti TCP/*Transfer Control Protocol*, data mentah *Vorbis* ini harus dibungkus dalam data *Ogg*. Data *Ogg* terdiri atas *page-page* yang didalamnya terdapat *segment-segment*, *header*, dan *lacing value*. Sebuah aliran data *Ogg* dapat hanya terdiri atas satu *page*.

Decoder *Ogg Vorbis*

Dari Gambar 8 pertama-tama data *Ogg* akan 'dibongkar' untuk mendapatkan data mentah *Vorbis*, yang kemudian akan dilakukan *decoding* data *floor* dan *residue*. Kemudian dilakukan *decoupling* kanal. Setelah itu data *floor* dan data *residue* dimasukkan dalam blok *dot product*. Keluaran dari *dot product* ini merupakan



Gambar 7 Diagram blok encoder Vorbis.
 (http://stofferke.port5.com/en_block/block_en.html)



Gambar 8 Diagram blok decoder Ogg Vorbis.

vektor spektrum audio. Vektor spektrum audio ini kemudian dijadikan masukan bagi IMDCT (*Inverse MDCT*) untuk mendapatkan data audio. Kemudian dilakukan penumpangan data keluaran IMDCT dengan blok sebelumnya untuk mendapatkan data lengkap. Pada proses penumpangan data ini, data blok aktif saat ini akan disimpan untuk kemudian ditumpangkan dengan data blok berikutnya. Setelah proses penumpangan ini selesai, maka diperoleh aliran data audio yang sesungguhnya.

KUALITAS SUARA Ogg Vorbis

Untuk menentukan kualitas suara hasil dari *codecs* audio, maka dapat dilakukan dua macam pengujian. Yang pertama adalah pengujian secara subyektif dan pengujian secara obyektif. Pengujian secara subyektif adalah pengujian berdasarkan hasil pendengaran kualitas suara yang dihasilkan oleh sebuah *codecs*. Pengujian ini lebih menitikberatkan pada subyektifitas pendengar. Kualitas secara subyektif inilah sebenarnya yang menjadi patokan utama bagi pemampatan data audio. Sedangkan pengujian secara obyektif adalah pengujian dengan cara melihat komponen obyektif dari data tersebut seperti *bit rate*, rasio pemampatan, dan rata-rata spektral yang dihasilkan. (www.xiph.org)

Test yang dilakukan oleh Fraunhofer dan Thomson, pengembang Mp3, diketahui bahwa untuk mendapatkan kualitas yang setara dengan kualitas CD audio, sebuah file Mp3 harus disandikan menggunakan *bit rate* sebesar 256 kbps dengan rasio pemampatan (*compression ratio*) sebesar 5:1.

Sementara itu *bit rate* sebesar 128 kbps menjadi standar. Meskipun kebanyakan orang dengan menggunakan perangkat audio yang baik dapat membedakan kualitas suara yang dihasilkannya dengan CD audio, namun kualitas suara yang dihasilkan masih terdengar cukup baik bagi kebanyakan orang. Dengan menggunakan *bit rate* 128 kbps, rasio pemampatan yang diperoleh adalah 10:1.

Ogg Vorbis tidak menekankan pada ukuran *bit rate* guna menentukan kualitas suara yang akan diperoleh (secara *default* Ogg Vorbis menggunakan VBR). Ogg Vorbis *encoder* secara normal tidak berdasarkan pada *bit rate* saja, namun juga berdasarkan pada faktor kualitas yang besarnya berkisar antara '-1' (terendah) hingga '10' (tertinggi) dengan kenaikan sebesar '0,01'. Sedangkan faktor kualitas yang normal berada diantara '0' hingga '10'. Faktor kualitas ini adalah sebuah ukuran tentang seberapa dekat terdengarnya sebuah *file* yang telah termampatkan dengan *file* aslinya. *Encoder* akan menggunakan bit sesuai dengan kebutuhan untuk memenuhi permintaan kualitas. Setiap pengaturan kualitas menghasilkan rata – rata *bit rate* kasar untuk sebuah potongan musik, namun ini dihasilkan dari bagaimana *encoder* diatur, *encoder* Vorbis sendiri tidak berpatokan pada suatu *bit rate* tertentu. Pengaturan *default* faktor kualitas pada *encoder Vorbis* adalah 3, di mana akan memberikan hasil yang cukup baik bagi rata – rata pengguna karena sudah memberikan kualitas suara yang lebih baik dari pada sebuah *file* Mp3 dengan *bit rate* sebesar 128 kbps, namun berukuran 10% lebih kecil.

Secara normal, sebuah file Ogg Vorbis akan menggunakan kopling kanal merugi, yang berarti kelebihan data antara kanal kiri dan kanal kanan akan digabungkan untuk menghemat ruang. Hal ini akan menjadikan *file* lebih kecil, namun juga berarti bahwa *image* stereo yang dihasilkan oleh sebuah *file* Ogg Vorbis tidak akan selalu identik dengan *image* stereo aslinya. Jika ini mengganggu, maka dengan menggunakan faktor kualitas 6 atau lebih kopling kanal merugi akan dimatikan dan semua kopling kanal akan menjadi tak merugi. Kebanyakan orang tidak dapat membedakannya, namun mungkin beberapa orang dapat membedakannya.

Penggunaan *variable bit rate* pada *file* yang digunakan untuk *streaming* pada web tidak dimungkinkan, karena *bit rate spike* dapat memperbesar kebutuhan *bandwidth*. Untuk kebutuhan semacam ini Ogg Vorbis

encoder dapat menggunakan *constant bit rate*. Selain itu juga tersedia pengaturan untuk nilai maksimal *bit rate* dan nilai minimal *bit rate*. Namun hal ini akan menghasilkan kualitas suara yang lebih rendah dibandingkan dengan pengaturan standar yang telah ada.

Perbandingan Codecs Ogg Vorbis dan Mp3

Untuk mengetahui sebaik apa *codecs* Ogg Vorbis, maka akan dibandingkan dengan kualitas Mp3 dan kualitas CD audio. Adapun langkah-langkah untuk mengetahui melakukan perbandingan adalah sebagai berikut.

- 1) CD audio di *rip* untuk mendapatkan file dengan format wav dengan kualitas CD audio.
- 2) File dengan format wav tersebut kemudian diubah menjadi format Ogg dengan menggunakan "oggenc" dengan faktor kualitas '5'. File format wav tersebut juga diubah menjadi format Mp3 dengan menggunakan "LAME" dengan menggunakan VBR dan faktor kualitas VBR sebesar '2'.

- 3) File dengan format wav, ogg dan Mp3 kemudian didengarkan dan dibandingkan untuk mengetahui kualitas suara secara subyektif.
- 4) Untuk mengetahui besarnya rata-rata *bitrate* dan *bit rate* yang digunakan pada tiap *frame* digunakan perangkat lunak "mp3stat".
- 5) Untuk mengetahui rata-rata spektral dari setiap format digunakan perangkat lunak "baudline", kemudian diperbandingkan ketiga format tersebut.

Hasil Perbandingan

File data audio yang akan dibandingkan dapat dilihat pada Tabel 1. Ketiga lagu tersebut diambil dari album kompilasi "image, emotional & relaxing" terbitan Sonny Music. Ketiga lagu tersebut adalah :

- 1) *Cheers to the Earth* yang dibawakan oleh Takefumi Haketa,
- 2) *Libertango* yang dibawakan oleh Yo-Yo-Ma,
- 3) *Rose* yang dibawakan oleh James Horner.

Tabel 1 File yang dibandingkan

Nama File	Format	Faktor kualitas	Jenis bit rate	Bit rate rata-rata (kbps)	Ukuran file (Mb)	Rasio ukuran file	Rasio pemampatan	durasi
Cheers To The Earth.wav	wav	-	-	1411,2	38,6			00:03:49
Cheers To The Earth.ogg	Ogg	5	VBR	169,19	4,6	11,92%	8,39 : 1	00:03:49
Cheers To The Earth.wav.mp3	Mp3	2	VBR	194,34	5,3	13,73%	7,28 : 1	00:03:49
Libertango.wav	wav	-	-	1411,2	31,9			00:03:09
Libertango.ogg	Ogg	5	VBR	151,51	3,4	10,66%	9,38 : 1	00:03:09
Libertango.wav.mp3	Mp3	2	VBR	174,42	3,9	12,23%	8,18 : 1	00:03:09
Rose.wav	wav	-	-	1411,2	29,1			00:02:52
Rose.ogg	Ogg	5	VBR	158,41	3,3	11,34%	8,82 : 1	00:02:52
Rose.wav.mp3	Mp3	2	VBR	155,5	3,2	11,00%	9,09 : 1	00:02:52

Tabel 2 Perbandingan bitrate pada format Ogg dan Mp3

	Cheers To The Earth		Libertango		Rose	
	Ogg	Mp3	Ogg	Mp3	Ogg	Mp3
Average bit rate	169,19 kbps	194,34 kbps	151,51 kbps	174,42 kbps	158,41 kbps	155,50 kbps
Durasi	229,20 detik	229,01 detik	189,67 detik	189,52 detik	172,89 detik	172,77 detik
Frame	12374	8776	11211	7263	7469	6621
32kbit:	0.7% (90)	0.9% (82)	0.2% (22)	0.3% (21)	0.4% (31)	0.5% (32)
40kbit:	0.0% (0)	0.0% (3)	0.0% (1)	0.0% (1)	0.0% (0)	0.0% (3)
48kbit:	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (2)	0.0% (0)	0.1% (4)
56kbit:	0.0% (0)	0.0% (2)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (1)
64kbit:	0.0% (0)	0.0% (1)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (3)
80kbit:	0.0% (1)	0.0% (4)	0.0% (0)	0.1% (9)	0.0% (0)	0.0% (2)
96kbit:	0.0% (0)	0.2% (18)	0.0% (0)	0.9% (66)	0.0% (1)	0.1% (7)
112kbit:	0.0% (3)	0.7% (58)	0.5% (61)	1.7% (122)	0.1% (11)	3.3% (220)
128kbit:	3.6% (449)	2.4% (209)	26.8% (2999)	5.5% (399)	3.3% (248)	23.7% (1572)
160kbit:	47.1% (5824)	30.2% (2653)	40.1% (4498)	43.9% (3185)	84.2% (6289)	61.2% (4050)
192kbit:	28.5% (3529)	39.6% (3479)	14.5% (1626)	41.0% (2979)	11.6% (867)	6.9% (457)
224kbit:	7.0% (864)	14.1% (1234)	7.6% (847)	4.7% (342)	0.1% (7)	2.3% (153)
256kbit:	6.3% (775)	5.7% (502)	10.0% (1120)	1.7% (122)	0.1% (4)	1.4% (96)
288kbit:	6.1% (756)	0.0% (0)	0.3% (37)	0.0% (0)	0.1% (9)	0.0% (0)
320kbit:	0.7% (83)	6.1% (531)	0.0% (0)	0.2% (15)	0.0% (2)	0.3% (21)

a) Perbandingan kualitas subyektif

Secara subyektif kualitas suara pada *file* audio dengan format *Ogg* dan Mp3 hampir sama, namun kualitas suara *file* audio dengan format *ogg* terasa lebih “bersih” sementara *file* dengan format mp3 sedikit terdengar adanya suara desis dibelakang yang cukup jelas, disamping itu format *Ogg* juga lebih terasa mendekati suara aslinya (format CD audio). Pemisahan stereo untuk kedua format sudah dirasa memuaskan, namun untuk perbedaan suara yang ekstrim (seperti pada lagu “*Cheers to the Earth*”) terada format *Ogg* terasa lebih baik.

Dengan melihat Tabel 1, tampak dengan ukuran *file* yang lebih kecil diperoleh kualitas subyektif yang lebih baik.

b) Perbandingan kualitas obyektif

Pada Tabel 1, dengan kualitas audio yang hampir sama, tampak bahwa rasio pemampatan *Ogg Vorbis* lebih tinggi dibandingkan dengan Mp3 kecuali pada lagu ketiga (*Rose.wav.Mp3*). Sementara perbandingan *bitrate* yang digunakan dapat dilihat pada Tabel 2, di mana terlihat Mp3 menggunakan *bit rate* rata-rata yang lebih besar dibandingkan dengan *Ogg*.

Pada pengamatan rata-rata spektral dengan menggunakan perangkat lunak *baudline* (grafik di lampiran), tampak secara keseluruhan, format *ogg* dapat memberikan keluaran yang hampir sama dengan format aslinya (CD audio). Pada format Mp3, rata-rata telah dilakukan pemotongan pada komponen frekuensi diatas 15 kHz. Sedangkan pada format *ogg* pemotongan baru terjadi disekitarnya 19 kHz. Pada grafik, tampak untuk lagu “*Rose*” pada format Mp3 telah dilakukan pemotongan pada 16 kHz. Hal ini menjelaskan ukuran *file* format Mp3 untuk lagu “*Rose*” lebih kecil dibandingkan dengan format *Ogg*.

PENUTUP

Ogg Vorbis sebagai suatu teknik pemampatan suara secara merugi yang masih baru memiliki beberapa keunggulan seperti sifatnya yang *free software*, sehingga memberikan kebebasan dalam penggunaan dan pengembangan oleh pihak lain.

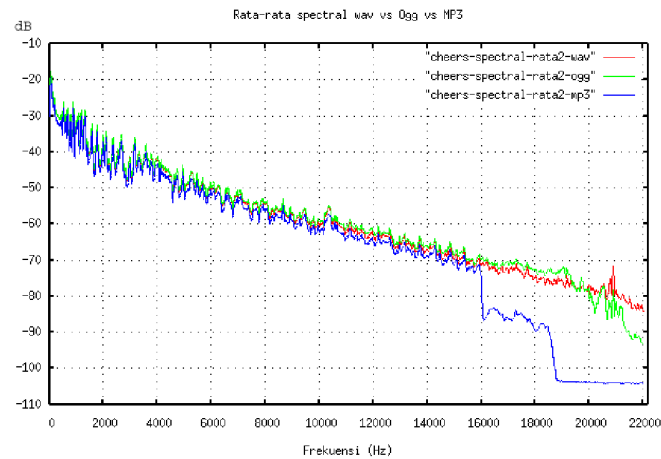
Secara umum dapat dikatakan bahwa kualitas suara yang dihasilkan oleh *codecs Ogg Vorbis* lebih baik dibandingkan dengan

Mp3. Suara yang dihasilkan oleh *Ogg Vorbis* lebih mendekati suara aslinya dibandingkan dengan Mp3. Selain itu dapat dikatakan juga bahwa untuk mendapatkan kualitas yang sama dengan kualitas Mp3, maka ukuran *file ogg* akan lebih kecil dibandingkan ukuran *file* Mp3.

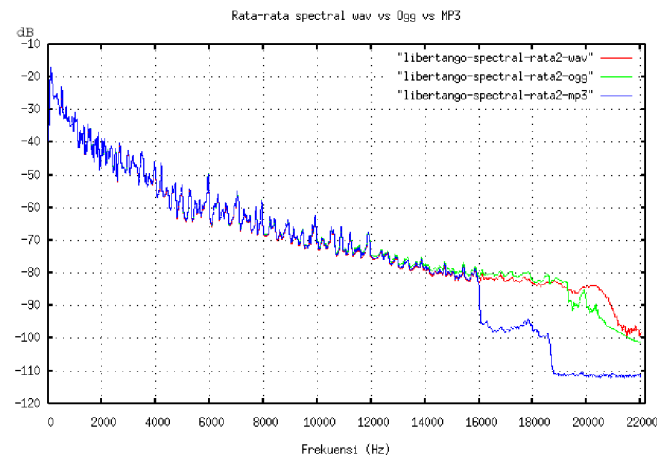
PUSTAKA

- _____, *Audio Compression*, <http://www.cs.sfu.ca/undergrad/CourseMaterials/CMPT479/material/notes/Chap4/Chap4.3/Chap4.3.html>, ____.
- _____, *Ogg Vorbis Documentation*, <http://www.xiph.org>, 2005.
- _____, *Ogg Vorbis Encoder Block Diagram*, http://stoffke.port5.com/en_block/block_en.html, 2005.
- _____, *Theory of Data Compression*, <http://www.data-compression.com/>, ____.
- _____, *Xiphopus mailing list archive*, <http://www.xiph.org/archives/>, 2003.
- Coleman Mike, *Vorbis Illuminated*, <http://www.mathdogs.com/vorbis-illuminated/>, 2001.
- Mitchell Graham, *An Introduction to compressed audio with Vorbis*, <http://grahammitchell.com/>, 2003.
- Mohammad H. Deghani dan Kuldip K. Paliwal, *Lossy and Lossless Coding and Compression of Encoded Audio Stream*, Microelectronis Engineering Research Conference, 2001.
- Pasi 'Albert' Ojala, *Compression Basics*, <http://www.cs.tut.fi/~albert>, ____.
- S. Pfeiffer, *RFC 3533 : The Ogg Encapsulation Format Version 0*, <http://community.roxen.com/developers/docs/rfc/rfc3533.html>, Mei 2003.
- Sporer Th., Brandenburg Kh., Edler B., *The use of multirate filterbanks for coding high quality digital audio*, 6th European Signal Processing Conference (EUSIPCO), Amsterdam, June 1992, Vol.1, halaman 211-214.
- Wright Keith, *Notes on Ogg Vorbis and the MDCT*, <http://www.free-shop.com>, 23 Mei 2003.

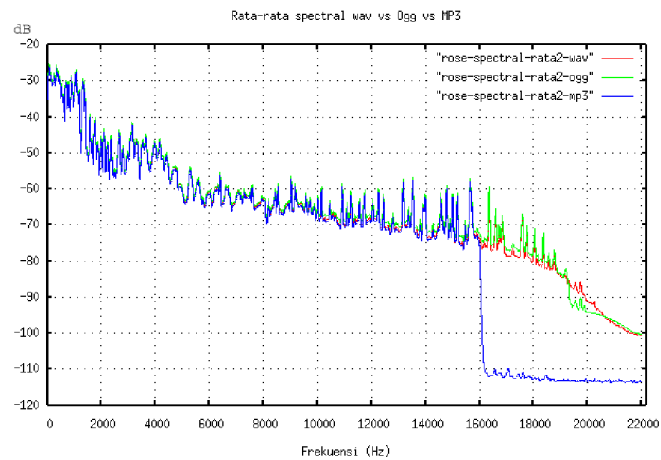
LAMPIRAN



Lampiran 1 Spektral rata-rata lagu "*Cheers to the Earth*".



Lampiran 2 Spektral rata-rata lagu "*Libertango*".



Lampiran 3 Spektral rata-rata lagu "*Rose*".